Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 237 (2024) 827–832

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry Sciences and Computer Science Innovation

Applying Blockchain Technology into Big Data: Advantages and Challenges

Hamed Taherdoost [a,*], Nazanin Moosavi [b,c], Nachaat Mohamed [d], Inam Ullah Khan [e]

[b] Department of Arts, Communications and Social Sciences, University Canada West, Vancouver, Canada
[b] QMinded | Quark Minded Technology Inc, Vancouver, Canada
[c] Research and Development Departement, Hamta Business Corporation, Vancouver, Canada
[e] Rabdan Academy, (HLS), Abu Dhabi, UAE
[e] Department of Electronic Engineering, School of Engineering and Applied Science, Isra University, Pakistan

Abstract

Big data and its analysis are crucial for other applications like the Internet of Things (IoT), data mining, cloud computing, etc. The volume of data is expanding along with the internet and other digital technologies. Traditional systems, however, cannot process this volume of data. Analyzing this much data will take a lot of work. However, there are several difficulties in assuring data security and privacy, such as safeguarding against data leakage and security breaches, flaws in open databases, and third-party data sharing, among others. This paper's goal is to analyze the benefits of blockchain in big data as well as its drawbacks and identify potential future approaches.

Keywords: Big Data; Blockchain; Internet of Things (IoT); Digital Technology

1. Introduction

Over the last ten years, data traffic has expanded at a worldwide level at an unprecedented pace, which accounts for the interest in "big data." By 2025, the big data market will be worth 229.4 billion dollars and will significantly lower costs across a range of industry verticals, including entertainment, manufacturing, transportation, and healthcare [1]. Despite the absence of a defined definition, several disciplines of research and engineering are investigating big data, including computer vision, the Internet of Things (IoT), and operation management. Big data refers to a new generation of technologies and architectures being researched for the purpose of analyzing vast amounts of data and capturing its key characteristics (e.g., analytics, knowledge discovery, and high velocity) [2]. Big data is viewed from a comparative perspective as extremely large and multidimensional dataset that cannot be managed, stored, analyzed, or captured using traditional database tools. Big data is defined from an architectural perspective as datasets with very high representation, velocity, and volume that require considerable horizontal scaling techniques for effective processing [3].

10.1016/j.procs.2024.05.171

However, there are several difficulties and problems with big data applications and methodologies, including big data intelligence, real-time data processing, data interpretation and management, computational scalability, and data privacy and security. Big data often incorporates many forms of sensitive information, such as age, address, and financial information, so security and privacy are significant considerations among these difficulties. For example, in Su et al. [4] study, a coalitional game and matching theory were jointly used to protect mobile social networks with massive data. This is only one example of the many methods and strategies that have been researched to safeguard data secrecy and private information. Wu et al. [5] looked at using reinforcement learning to create a security-aware algorithm for a smart grid system.

As a ledger technology, blockchain has recently become one of the most alluring options for ensuring privacy and security in large data systems [6]. For industrial IoT applications, blockchain can offer high-quality data and secure data sharing [7]. A blockchain-based approach for securing data collecting in ad hoc networks and rewarding mobile nodes for effective data collection was put out in [8]. In addition, blockchain and edge computing were employed to enhance data quality and manage the computationally intensive activities required by IoT devices while ensuring security [9, 10].

By supplying effective network management and security features capabilities for enabling freshly developing big data services and applications, blockchain has a significant potential to revolutionize existing big data systems thanks to its distinct benefits [11]. This survey outlines the advantages of blockchain in big data as well as its challenges and suggests possible directions for the future.

2. Advantages of Employment of Blockchain Technology in Big Data

Big data and blockchain technologies have developed incrementally alongside human civilization [10]. Although big data technology is evolving, there are still unresolved problems including data silos, low-quality data, and data leakage, and big data risk management is not yet ideal [12]. The fundamental ideas of Blockchain are decentralization, accessibility, and transparency. Blockchain can address the problem of online trust, hence boosting the expansion of big data and the digital economy [13]. Once the massive data of the Internet met the blockchain, it was supported by a high level of confidence, which promoted the formation of new data firms. Some typical big data social life situations may be integrated with blockchain technology by employing smart contracts to create a new wave of apps. They will concentrate on enhancing accountability, openness, and trust while streamlining corporate procedures and legal restraints. This part will examine some possible advantages of merging blockchain with big data. Fig. 1 summarises the advantages of using blockchain technology in big data.

2.1. Reducing Cost

Blockchain technology has the potential to lower business expenses. With anti-tampering and anti-counterfeiting properties, blockchain technology addresses the trust and reliable value transfer problems at a very low cost and can create a more shared, transparent, open, trustworthy, and verifiable dependable system. It's crucial to assess the cost savings brought on by using blockchain in current systems [14]. Blockchain can save expenses by lowering transaction and processing costs, particularly in banks. Blockchain, according to Hassani et al. [15], is upending the financial sector and causing a rise in big data use.

2.2. Data Sharing

Data security problems in the age of big data include not only the protection of individual privacy but also the issue of data analysis, which aims to predict people's status and behavior. Further analysis is made possible by the organized and full nature of the data kept on the blockchain. Blockchain not only increases the size of big data but also makes it safer and more valuable since blockchain large data is organized and ready for big data analytics [16]. Blockchain may increase data integrity and assure confidence between various stakeholders in the cross-border logistics network, as shown by Huang [17]. Transaction data, financial contract data, no de-aware data, location data, and order data will all be combined in the cross-border supply chain in a distributed way, and all parties will agree on the amount of detail

of the sharing. These real-time, traceable, and better-integrated data may be used by a data analyst for real-time optimization and analysis to create a global optimization method.

2.3. Data Analysis

The blockchain has partially fixed the security issue around data exchange. Using blockchain technology, big data acquired through data analysis may be stored in the blockchain network. The project team will not mistakenly reuse previously employed data or conduct the same data analysis as other teams. Blockchain technology may also assist academics in monetizing their study via the analysis of recorded transactions. The normal multi-center situation in healthcare is that no one institution has access to all the data. Various healthcare organizations want to gather health information via medical data sharing and exchange as the need for personal health control rises [18]. Cichosz et al. [19] proposed a concept for a blockchain-based system for sharing Diabetes Health Care Data. This approach places patients at the focus of controlling their data, privacy concerns, and data exchange.

2.4. Transparency and Credibility

Big data can offer data analysis for demanders via the automated execution of smart contracts using blockchain as an intermediary. By using smart contracts, it minimizes redundant human involvement. Through accurate analysis and scanning of all data, the blockchain network is linked to the automated execution of smart contracts, which is unknown but trusted by many parties. Blockchain and big data may work together to endorse transactions, secure data mutual trust, and give a guarantee for the final design of value exchanges and asset exchange in the realm of the digital economy. Big data and blockchain have the potential to increase industrial transparency. For instance, several illicit enterprises are vying for clients' money in the fields of vehicle insurance and auto financing. Consumer rights and interests may be successfully protected by increasing industry openness. Maesa et al. [20] proposed a blockchain-based technique for the development of auditable access control systems. Due to publicly auditable evidence of wrongdoing, both the resource owner and the individual requesting access may rapidly notice improper authorization or denial of access.

2.5. Privacy and Security

Through its decentralized architecture, blockchain technology assures the security and anonymity of huge data. The fact that the majority of the data is kept on centralized servers increases the risk of data loss and leaking. Attackers on the internet often target them. Due to the decentralized nature of blockchain technology, it is difficult for thieves to access and change data on a large scale. Additionally, the blockchain's transaction data is open and accessible, including the transaction address, transaction amount, and transaction time. However, the transaction address's owner's identity is unknown. The identity of the user and user data may be segregated using the encryption capabilities of blockchain technology. Healthcare is where the usage of blockchain and big data is most prevalent. The integrity, privacy, and security of medical data must be upheld by the healthcare sector. A secure and reliable data management system is thus required. Using a permissioned blockchain network, Azaria et al. [21] launched MedRec, which provides patients with a thorough, immutable record and simple access to their health information across providers and treatment venues. MedRec maintains authentication, secrecy, and accountability by using blockchain—important factors to take into mind while managing sensitive information. The blockchain supports differential privacy because it can execute asymmetric encryption. Patients may exchange health records in this manner without revealing any private information.

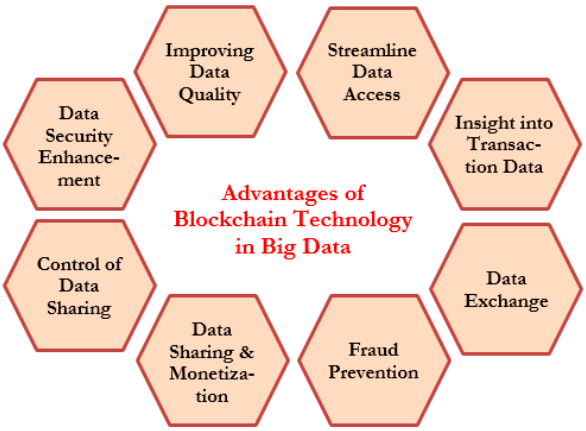

Fig. 1. A list of blockchain's benefits for big data

3. Challenges and Future Perspectives

Businesses might easily fall victim to the buzz around blockchain technologies without being aware of their existing practical limits. Particularly in settings where centralization and transactional malleability are crucial, the blockchain is not the best substitute. One such instance is the situation when businesses still place a high priority on data input. Human error-related faults will always exist on the blockchain, necessitating the creation of new authorizations solely to monitor these rights [22].

The enormous number of possible applications and the technologies they depend on, as well as the enabling technology for blockchain, will all continue to develop over time. Consumers will get greater information and assistance in cultivating a community devoted to creating such blockchain protocols as a result of this future expansion [23]. Unfortunately, that day is still quite some ways off. The difficulties in integrating big data and blockchain are summarized in Table 1. Significant blockchain technologies are advancing quickly every day, bringing forth new improvements in the big data analytics industry. Analytics must advance as more individuals begin to use blockchain-driven services to provide high-quality insights from the available data. Big data demands data processing systems with enhanced computing capabilities due to its unique unpredictability, volume, veracity, value, and complexity.

Table 1. Integration issues with blockchain and big data

Application	Issues	Source
Cyber-physical social infrastructures	The computational burden of blockchain	[24]
Big data to optimize mobile network performance	Big data's complexity	[25]
Early standardization for the immutability of blockchain	Standardization of blockchain	[26]
Big data trade and smart contract	Privacy and security in big data interchange	[27]
Big data blockchain smart contracts	Enhancing privacy and security for big data	[28]

Additionally, the decentralized and distributed ledger blockchain enables safe, transparent, and unchangeable data transfers, but these services need additional processing power to be successful. The system performs poorly as a result of the unforeseen computational complexity that results from the integration of these sophisticated large data with blockchain [29]. To reduce the computing resource used for blockchain, adaptable blockchain architectures should be selected. This will allow a 5G network connection to be used for speedier services.

4. Conclusion

Two emerging technologies that are high on the agendas of many firms are blockchain and big data. Both are anticipated to significantly affect how companies and organizations run in the next years. Although big data has been around for a while, blockchain technology is presently experiencing a surge in popularity. In the field of IT, big data and its analysis are crucial for other applications like IoT, data mining, and cloud computing, among others.

The volume of data is expanding along with the internet and other digital technologies. Traditional systems, however, cannot process this volume of data. Analyzing this much data will take a lot of work. However, there are several difficulties in assuring data security and privacy, such as safeguarding against data leakage and security breaches, flaws in open databases, and third-party data sharing, among others.

This study covered the benefits of blockchain for big data, including cost savings, data sharing and analysis, openness and credibility, privacy and security, and cost reduction. In addition, difficulties and orientations for the future have been discussed. More advancement in the interaction between big data analytics and blockchain may be anticipated as this industry's developments continue. More specific use cases for data analysis and big data management will be developed and studied as technology advances and developments take place.

References

[1] Adadi, A., A survey on data-efficient algorithms in big data era. Journal of Big Data, 2021. 8(1): p. 1-54.
[2] Gantz, J. and D. Reinsel, Extracting value from chaos. IDC iview, 2011. 1142(2011): p. 1-12.
[3] Pouyanfar, S., et al., Multimedia big data analytics: A survey. ACM computing surveys (CSUR), 2018. 51(1): p. 1-34.
[4] Su, Z. and Q. Xu, Security-aware resource allocation for mobile social big data: A matching-coalitional game solution. IEEE transactions on big data, 2017. 7(4): p. 632-642.
[5] Wu, J., et al., Big data analysis-based security situational awareness for smart grid. IEEE Transactions on Big Data, 2016. 4(3): p. 408-417.
[6] Taherdoost, H., A Critical Review of Blockchain Acceptance Models—Blockchain Technology Adoption Frameworks and Applications. Computers, 2022. 11(2): p. 24.
[7] Liu, C.H., Q. Lin, and S. Wen, Blockchain-enabled data collection and sharing for industrial IoT with deep reinforcement learning. IEEE Transactions on Industrial Informatics, 2018. 15(6): p. 3516-3526.
[8] Liu, G., et al., B4SDC: A blockchain system for security data collection in MANETs. IEEE Transactions on Big Data, 2020.
[9] Xu, X., et al., BeCome: Blockchain-enabled computation offloading for IoT in mobile edge computing. IEEE Transactions on Industrial Informatics, 2019. 16(6): p. 4187-4195.
[10] Moosavi, N. and H. Taherdoost, Blockchain and Internet of Things (IoT): A Disruptive Integration, in 2nd International Conference on Emerging Technologies andIntelligent Systems (ICETIS 2022). 2022, Lecture Notes in Networks and Systems, Springer: virtual conference.
[11] Taherdoost, H., Blockchain Technology and Artificial Intelligence Together: A Critical Review on Applications. Applied Sciences, 2022. 12(24): p. 12948.
[12] Moosavi, N. and H. Taherdoost. Blockchain-Enabled Network for 6G Wireless Communication Systems. in International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI 2022). 2022. Coimbatore, India: Engineering Cyber-Physical Systems and Critical Infrastructures, Springer.
[13] Sharma, J. and H. Taherdoost, Impact of Blockchain Technology on the Development of E-Businesses, in Advances in Data Computing, Communication and Security. 2022, Springer. p. 391-396.
[14] Sawa, T., Blockchain technology outline and its application to field of power and energy system. Electrical Engineering in Japan, 2019. 206(2): p. 11-15.
[15] Hassani, H., X. Huang, and E. Silva, Banking with blockchain-ed big data. Journal of Management Analytics, 2018. 5(4): p. 256-275.
[16] Fedak, V., Blockchain and big data: The match made in heavens. Towards Data Science, 2018.
[17] Huang, Z. From Data Science to Blockchain–Analytics in Cross-Border Logistics [w:]. in Proceedings of the 8th SIGKDD International Workshop on Urban Computing (UrbComp'19). 2019.
[18] Chen, J., Z. Lv, and H. Song, Design of personnel big data management system based on blockchain. Future generation computer systems, 2019. 101: p. 1122-1129.
[19] Cichosz, S.L., et al., How to use blockchain for diabetes health care data and access management: an operational concept. Journal of diabetes science and technology, 2019. 13(2): p. 248-253.
[20] Maesa, D.D.F., P. Mori, and L. Ricci, A blockchain based approach for the definition of auditable access control systems. Computers & Security, 2019. 84: p. 93-119.
[21] Azaria, A., et al. Medrec: Using blockchain for medical data access and permission management. in 2016 2nd international conference on open and big data (OBD). 2016. IEEE.
[22] Taherdoost, Hamed. Smart Contracts in Blockchain Technology: A Critical Review. Information, 2023, 14.2: 117.
[23] Taherdoost, Hamed; MADANCHIAN, Mitra. Blockchain-Based New Business Models: A Systematic Review. Electronics, 2023, 12.6: 1479.

[24] Tan, L., et al., A blockchain-based access control framework for cyber-physical-social system big data. IEEE Access, 2020. 8: p. 77215-77226.
[25] Zheng, K., et al., Big data-driven optimization for mobile networks toward 5G. IEEE network, 2016. 30(1): p. 44-51.
[26] Hofmann, F., et al. The immutability concept of blockchains and benefits of early standardization. in 2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K). 2017. IEEE.
[27] Chen, Y., et al., FaDe: a blockchain-based fair data exchange scheme for big data sharing. Future Internet, 2019. 11(11): p. 225.
[28] Liu, J. and Z. Liu, A survey on security verification of blockchain smart contracts. IEEE Access, 2019. 7: p. 77894-77904.
[29] Taherdoost, Hamed. Blockchain: A catalyst in fintech future revolution. Future Technology (FUTECH), 2023.